

Suchmaschinen: Status quo und Entwicklungstendenzen

ANDRE SKUSA

andre.skusa@lycos-europe.com

CHRISTIAN MAASS

christian@ebusiness-management.info

Suchmaschinen erfassen und sortieren die Inhalte des Internets und sind über die letzten Jahre hinweg zu den meistgenutzten Webanwendungen avanciert. Im Zuge der Auseinandersetzung mit dem Suchmaschinenmarkt stehen dabei häufig Unternehmen wie Google, Yahoo! oder Microsoft im Vordergrund. In jüngerer Zeit lässt sich jedoch beobachten, dass neben diesen algorithmenbasierten Suchmaschinen immer häufiger alternative Suchdienste in den Markt treten, wie z. B. soziale, semantische oder so genannte konvergente Suchdienste. In diesem Beitrag wird vor diesem Hintergrund aufgezeigt, dass diese Suchdienste zum jetzigen Zeitpunkt noch nicht mit Google und anderen führenden Suchmaschinen konkurrieren können. Es zeigt sich jedoch, dass eine Kombination verschiedener Suchdienste zu einer Verbesserung der Ergebnisqualität von Suchmaschinen beitragen kann.

1. EINLEITUNG¹

Mit dem exponentiellen Wachstum des Internets sind Suchmaschinen wie Google und Yahoo! zu den meistgenutzten Internetanwendungen avanciert, die von etwa 90 Prozent der Internetnutzer in Deutschland regelmäßig zu Recherchezwecken genutzt werden. Insbesondere im Vorfeld von Kaufentscheidungen spielen sie – sowohl für Unternehmen als auch für Privathaushalte – eine immer wichtigere Rolle, um gezielt nach bestimmten Produkten und Dienstleistungen zu recherchieren. Zur Befriedigung ihrer Informationsbedürfnisse können die Anwender grundsätzlich auf eine Vielzahl verschiedener Suchmaschinen zurückgreifen; allein für den deutschsprachigen Raum sind weit über hundert verschiedene Suchmaschinen verfügbar (vgl. für einen Überblick z. B. www.mysuche.de). Mit Blick auf die Praxis muss jedoch ein Oligopol algorithmenbasierter Suchmaschinen – bestehend aus Google, Yahoo! und MSN – konstatiert werden, das den weltweiten Markt dominiert; unter algorithmischer Suche versteht man dabei ein automatisiertes Verfahren, das im Internet vorhandene Seiten erfasst, mit Hilfe statistischer Methoden bewertet, sortiert und zu den Worten der Suchanfrage in Beziehung setzt.

Die Vormachstellung von diesen Suchmaschinen wird in der Öffentlichkeit bereits seit geraumer Zeit kritisiert. Teilweise wird in diesem Zusammenhang darauf hingewiesen, dass aufgrund dieser Marktkonstellation eine Gefahr für die Informationsvielfalt und -auswahl im Internet besteht, die nahezu durch eine Handvoll

¹ Dieser Artikel entstand unter Mithilfe von Jan Hoenck, Jan van Hamme und Rembert Wohler.

2 Status quo und Entwicklungstendenzen

Unternehmen mit kommerziellen Interessen beeinflusst werden [1]. Zu ähnlichen Äußerungen kommt es seitens der Politik [2, 3], die bereits seit längerer Zeit auf die Monopolisierungstendenzen im Suchmaschinenmarkt hinweist und sich für ein größeres Angebot alternativer Suchdienste ausspricht. In jüngerer Zeit kann man tatsächlich beobachten, dass in zunehmendem Maße Suchmaschinen mit alternativen Suchansätzen in den Markt treten. Möglichkeiten alternativer Suchansätze bestehen beispielsweise darin, das Wissen menschlicher Nutzer in den Suchprozess einzubeziehen (soziale Suchdienste) oder die bisherigen algorithmenbasierten Verfahren durch semantische Technologien zu verbessern (semantische Suchdienste). Weiterhin sind auch Kombinationen der hier genannten Suchansätze denkbar (konvergente Suchdienste), wie z. B. die Bewertung und Korrektur automatisch erzeugter Suchergebnisse durch Benutzer. Exemplarisch hierfür sei die Suchmaschine Wikia genannt; Abbildung 1.1 zeigt die hier angesprochenen Suchdienste im Überblick.

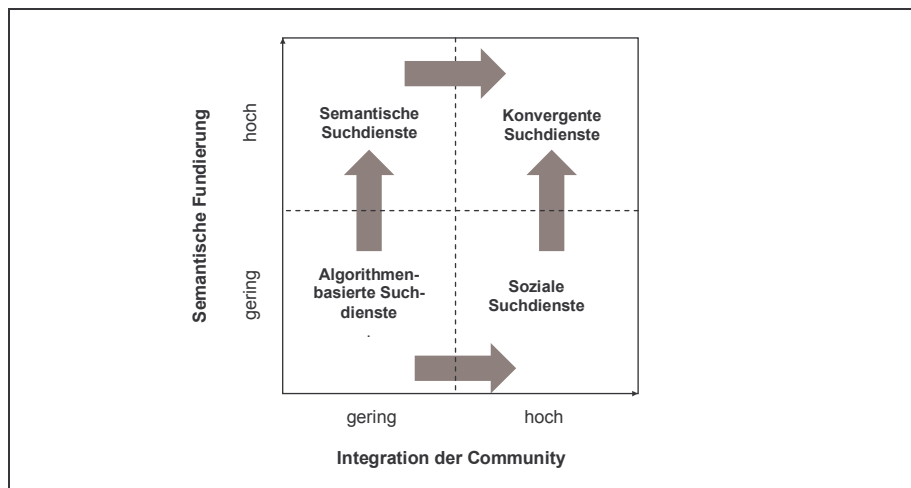


Abb. 1.1 Einteilung unterschiedlicher Typen von Suchdiensten

In Anbetracht der hier skizzierten Entwicklungstendenzen drängt sich unweigerlich die Frage auf, welche Bedeutung alternative Suchdienste im Allgemeinen und die aus der Kombination resultierenden konvergenten Suchdienste im Speziellen haben. Im weiteren Verlauf dieses Beitrags wird diese Frage ausführlicher beleuchtet. Um für die Auseinandersetzung mit diesem Thema zu sensibilisieren, gilt es zunächst jedoch, die Entwicklung und wirtschaftliche Bedeutung des Suchmaschinenmarktes aufzuarbeiten und die in diesem Kontext diskutierten algorithmischen Suchmaschinen zu skizzieren. Auf dieser Grundlage erfolgt anschließend eine Auseinandersetzung mit alternativen Suchdiensten. Dabei wird aufgezeigt, dass in der Verknüpfung von algorithmenbasierten und sozialen Suchdiensten ein interessanter, aber weitestgehend unerforschter Ansatz zu sehen ist, um die Qualität der Suchmaschinenergebnisse zu verbessern. Damit wird gleichzeitig die konzeptionelle Grundlage für den weiteren Verlauf dieses Buches gelegt.

2. ENTWICKLUNG UND WIRTSCHAFTLICHE BEDEUTUNG ALGORITHMISCHER SUCHMASCHINEN

Die Entwicklung des Suchmaschinenmarktes ist vergleichsweise jung und eng mit der Entwicklung des Internets und dessen wichtigstem Dienst, dem *World Wide Web* (WWW) verbunden. Der Grundstein des WWW wurde im Jahr 1989 am schweizerischen Kernforschungsinstitut CERN gelegt. Hier arbeitete man daran, Dokumente elektronisch miteinander zu verknüpfen, um auf diesem Wege den Informationsaustausch zwischen den Wissenschaftlern zu forcieren. Die Zahl dieser so genannten Hypertexte war bereits nach kurzer Zeit für Menschen nicht mehr überschaubar und es entstand ein Bedarf nach Hilfsmitteln, um die Hypertexte zu durchsuchen. Nachdem mit Archie bereits im Jahr 1990 ein Suchroboter für FTP-Dateien zur Verfügung stand, entstand im Frühjahr 1993 der erste Suchroboter für das WWW, der World Wide Web Wanderer. Er durchsuchte zwischen 1993 und 1996 zwei Mal pro Jahr das weltweite Datennetz [4]. Mit der Verbreitung solcher Suchmaschinen wurden die Verzeichnisdienste zunehmend abgelöst, in denen mittels manueller Zuordnung das WWW katalogisiert wurde; die heute noch bekannte Suchmaschine Yahoo! startete 1995 als ein solcher Verzeichnisdienst.

Ab Mitte der 90er Jahren entstanden mit dem zunehmenden Wachstum des Internets vor allem im universitären Umfeld weitere Suchmaschinen. Der im Jahr 1994 entwickelte *RBSE Spider* und der an der Universität von Washington entstandene Webcrawler waren z. B. die ersten algorithmischen Suchtechnologien, die ihre Ergebnisse bewertet haben (Ranking) und sie nach dieser Bewertung sortierten. Etwa zeitgleich entstand an der Carnegie Mellon University die Suchtechnologie Lycos, die in ihrer Relevanzbewertung auch die „word proximity“ berücksichtigte – also die Nähe verschiedener Suchwörter zueinander – und damit schnell Marktanteile gewinnen konnte.

Den bislang skizzierten algorithmischen Suchmaschinen ist gemein, dass sie die im Web auffindbaren Daten mit Hilfe von Crawlern sammeln, automatisch analysieren und dem Anwender den Zugriff auf diesen Index in Form einer Webseite ermöglichen. Nach diesem Prinzip funktioniert grundsätzlich auch die 1998 gegründete Suchmaschine Google, die diese Aufgaben jedoch wesentlich besser als ihre Konkurrenten bewältigte. Genauer gesagt gelang Google ein Durchbruch bei der Indizierung der im Web auffindbaren Dokumente. Während die seinerzeit etablierten Suchmaschinen wie Lycos und AltaVista die Relevanz ihrer Ergebnisse hauptsächlich anhand der im Dokument auffindbaren Schlüsselwörter bestimmten, integrierte Google zusätzlich den so genannten PageRank, nach dem die Relevanz eines Dokumentes rekursiv von den Verweisen auf dieses Dokument abhängt. Das bedeutet, dass eine Seite umso höher bewertet wird, je mehr andere Seiten auf sie verweisen, d.h. je bekannter sie ist. Diesem Ansatz liegt die Annahme zugrunde, dass auf Webseiten mit qualitativ guten Inhalten wesentlich mehr Links verweisen als auf inferiore Webangebote. Nicht zuletzt dieses Verfahren verhalf Google – trotz des relativ späten Markteintritts 1998 – zum Durchbruch und dem bis heute stetig wachsenden Marktanteils. Er beträgt in den USA etwa 61 Prozent, während Yahoo! und MSN mit 20,4 Prozent und 9,1 Prozent deutlich abgeschlagen sind (vgl. Abb. 1.2). Einige einst bekannte Suchmaschinen wie Alta Vista haben nahezu keine Marktbedeutung mehr. Noch drastischer fällt die Vormachstellung von Google

4 Status quo und Entwicklungstendenzen

allerdings in Deutschland aus, wo das Unternehmen einen Marktanteil von rund 90 Prozent aufweisen kann.

	Marktanteil im April 2008	Anzahl der Suchanfragen im April 2008 (in Millionen)
Google	61,6 %	6.514
Yahoo!	20,4 %	2.159
Microsoft	9,1 %	961
AOL	4,6 %	491
Ask.com	4,3 %	458

Abb 1.2 Marktanteile im deutschen Suchmaschinenmarkt [5]

2.1 Wirtschaftliche Bedeutung

Sind Werbeeinnahmen in der Internetbranche generell eine der Haupteinnahmequellen, so trifft dies auf Suchmaschinen in noch stärkerem Maße zu. So zählt die Suchmaschinenwerbung gegenwärtig zu den beliebtesten Werbeformen im Internet und weist im Vergleich zu traditionellen Werbeformen überdurchschnittliche Wachstumsraten auf. Allein in Deutschland wurden im Jahr 2007 mit dieser Werbeform knapp 1,2 Milliarden Euro und damit fast 40 Prozent mehr als im Vorjahr umgesetzt (vgl. auch Abb. 1.3). Mittelfristig rechnet man sogar damit, dass die Suchmaschinenwerbung – ähnlich wie in den USA – deutlich mehr als 50 Prozent der gesamten Online-Werbeausgaben auf sich vereinigt; die Schätzungen bezüglich des Volumens des weltweiten Online-Werbemarktes liegen zwischen 20 und 50 Milliarden Euro [6, 7]. Insofern ist auch in Zukunft damit zu rechnen, dass algorithmische Suchmaschinen weiterhin hohe Gewinne erwirtschaften.

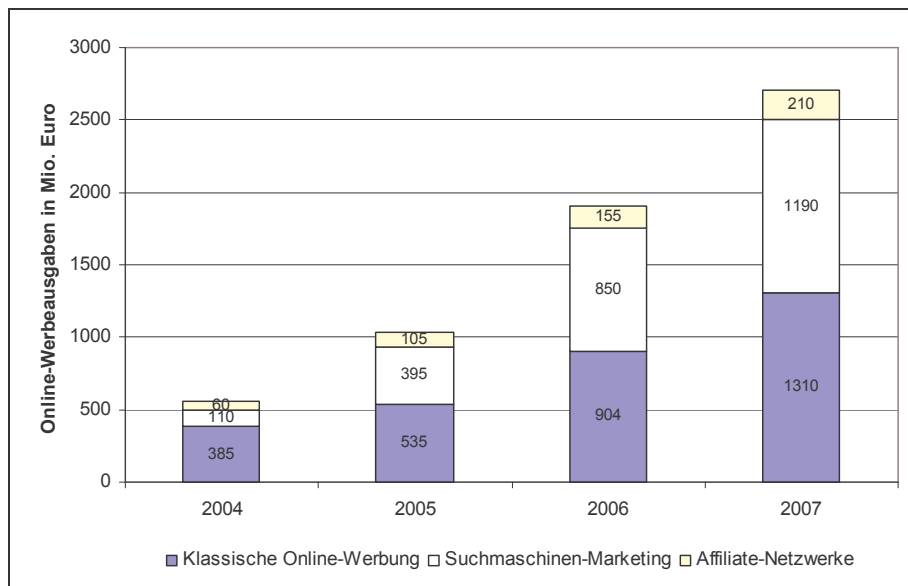


Abb. 1.3 Entwicklung des deutschen Online-Werbemarktes [8]

Die Funktionsweise des eigentlichen Erlösmodells von Suchmaschinen basiert vor diesem Hintergrund auf einem relativ einfachen Prinzip: Die Werbetreibenden müssen zur Platzierung ihrer Werbeanzeigen lediglich bestimmte Schlüssel- bzw. Suchwörter festlegen, unter denen die Anzeigen erscheinen sollen. Die Anzeigen als solche verweisen dann auf deren Internetpräsenzen, wobei für jeden Klick auf die Anzeige ein bestimmter Preis an den Suchmaschinenbetreiber zu entrichten ist; diese Form der Abrechnung wird im Allgemeinen als „Pay per Click“ bezeichnet. Bei Suchmaschinenwerbung werden dem Anwender folglich nur die zu seiner Suchanfrage passenden Anzeigen präsentiert. Dadurch steigt im Vergleich zu unspezifischen Werbeeinblendungen die Wahrscheinlichkeit, dass der Anwender auf die Anzeige reagiert.

Auf Basis des hier skizzierten Erlösmodells erwirtschaften die etablierten Suchmaschinen Gewinne in Milliardenhöhen. Allen voran ist hier das Unternehmen Google zu nennen, dessen Erlösmodell zu 99 Prozent auf so einer Werbefinanzierung basiert, womit jedes Quartal ein Gewinn von etwa 1 Milliarde US-Dollar einhergeht [9]. Die hohe Akzeptanz von Google liegt vor allem darin begründet, dass neben der neuartigen Bewertung der Suchergebnisse (PageRank, s. o.) und dem puristischen, auf das notwendigste beschränkten Design, die eingeblendete Werbung zu den Suchergebnissen passend und unauffällig ist. So kann diese Zusatzinformation zwar genutzt werden, ist aber trotzdem eindeutig als Werbung erkennbar. So wird der Nutzer nicht vom eigentlichen Suchergebnis abgelenkt, der Betreiber aber kann aufgrund der Beliebtheit der Seite genug aus den Werbeeinnahmen Erlösen.

Aufgrund der hier skizzierten wirtschaftlichen Potenziale erstaunt es kaum, dass alternative Suchdienste immer stärker in den Markt drängen, um vom lukrativen Geschäft mit der Suchmaschinenwerbung zu profitieren. Sowohl in der Wirtschaftspresse als auch in der Wissenschaft rückt dabei die Auseinandersetzung mit

- sozialen Suchdiensten
- semantischen Suchdiensten und
- konvergenten Suchdiensten

immer mehr in den Vordergrund. Diese Ansätze werden im weiteren Verlauf ausführlicher beleuchtet.

3. CHARAKTERISIERUNG UND MARKTÜBERBLICK ALTERNATIVER SUCHDIENSTE

Neben algorithmischen Suchmaschinen erfreuen sich in jüngerer Zeit auch so genannte soziale Suchdienste zunehmender Beliebtheit. Dabei handelt es sich um spezifische Ausprägungsformen sozialer Softwarelösungen, die in einer allgemeinen Begriffsumschreibung auf den Aufbau von sozialen Netzwerken sowie die Publikation und Verteilung von Informationen innerhalb eines Netzwerkes abstellen. „Social-Software-Systeme sind [...] umfassende sozio-technische Systeme, die auf Basis technischer und sozialer Vernetzung durch einfach zu bedienende Informationssysteme gemeinsam in einem bestimmten Themenfeld Leistungen

generieren“ [10]. Bei sozialen Suchdiensten bestehen diese Leistungen darin, dass die Informationsbedürfnisse der Anwender nicht von Suchmaschinenrobotern, sondern von einer Community aus Freiwilligen erbracht werden, die sich in erster Linie aus altruistischen Motiven bei diesen Diensten einbringen. Exemplarisch für solche Suchdienste wird im Folgenden auf

- soziale Bookmarksysteme und
- Frage-Antwort-Dienste eingegangen.

3.1 Soziale Bookmarkdienste

Soziale Bookmarkdienste wie z. B. del.icio.us, Furl oder Mister Wong – werden teilweise bereits als Alternative für die algorithmischen Suchmaschinen. gehandelt [11]. Diese Dienste bieten den Benutzern die Möglichkeit ihre persönlichen Lesezeichen (Bookmarks) im Netz zu speichern, zu verwalten und anderen Nutzern zur Verfügung zu stellen. Einmal gespeichert, können diese gern besuchten, nützlichen und persönlich ausgesuchten Webseiten jederzeit wieder gefunden und von vielen genutzt werden. Dies ermöglicht eine Suche in einer von menschlichen Anwendern zusammengestellten Treffermenge anstatt in Ergebnissen eines statistischen Bewertungsverfahrens. Die Funktionsweise dieser Dienste beruht dabei im Wesentlichen auf Tags [12]. Hierbei handelt es sich um frei wählbare Stichworte, die den Inhalt der gespeicherten Seite möglichst treffend charakterisieren sollen. Die zugeordneten Tags werden genutzt, um die auf den Seiten enthaltene Information zu beschreiben und auffindbar zu machen. Durch diese Tags entsteht im Laufe der Zeit eine Folksonomy [12], also eine durch eine große Menge von Personen definierte Klassifikation von Internetadressen. Anders formuliert können auf diesem Wege inhaltlich wertvolle, aber statistisch irrelevante Informationen gefunden werden, sofern sie von anderen Benutzern in den gemeinschaftlich gepflegten Index aufgenommen und sinnvoll annotiert werden. Dies führt unmittelbar zu der Frage, inwieweit eine Kombination dieser Suchdienste mit algorithmenbasierten Suchmaschinen zu einer Verbesserung der Ergebnisqualität beitragen könnte (vgl. hierzu insbesondere den Beitrag von Gammer u. a. in diesem Band).

Die gegenwärtig am Markt existierenden Bookmarksysteme – wie z. B. Mister Wong, del.icio.us, Furl, OneView, Bibsonomy, CiteULik, LinkArena usw.. – unterscheiden sich vor diesem Hintergrund nicht gravierend voneinander und haben ähnliche Features. Ebenso arbeiten alle Dienste mit freiem Tagging und bieten die Möglichkeit, bereits bestehende Lesezeichensammlungen zu importieren und die beim Dienst gespeicherten Lesezeichen zu exportieren. RSS-Feeds und die Integration in andere Tools können bei jedem Anbieter in Anspruch genommen werden. Einige Anbieter stellen ihren Anwendern jedoch besondere Dienste zur Verfügung: So werden bei Furl z. B. Kopien ganzer Webseiten gespeichert. Mit Oneview lassen sich hingegen auch Artikel und Videos verwalten. Ähnlich funktioniert auch Linkarena, wo auf Basis eines Datenmanagers Dokumente und Bilder hochgeladen und verwaltet werden können.

In der Praxis spielen diese Dienste im Hinblick auf ihre Reichweite im Vergleich zu Google. zum gegenwärtigen Zeitpunkt allerdings noch keine vergleichbare Rolle. Das liegt unter anderem daran, dass es sich bei der Vergabe von Tags um einen sehr

fehleranfälligen Prozess handelt. So kommt es unter anderem durch Rechtschreibfehler, fehlende Synonymerkennung (und dadurch verursachtes unkontrolliertes Wachstum der Menge der Tags) und Kombinationen unsachlich zugeordneter Tags nahezu zwangsläufig zu einer Verwässerung der Suchergebnisse. Es kann zwar etwas Ordnung in Tag-Sammlungen gebracht werden, indem den Benutzern Vorschläge beim Tagging unterbreitet werden. Fehlerhafte Zuordnungen werden hierdurch jedoch nicht verhindert. Als problematisch ist weiterhin die nur schwer zu gewährleistende Aktualität solcher Dienste anzusehen – z. B. bei der Nachrichtensuche –, da sie in der Regel auf eine längerfristige Speicherung von wichtigen Bookmarks ausgelegt sind. Aufgrund der kurzen Halbwertszeit von Nachrichten werden diese jedoch nur selten in die private Bookmarksammlung aufgenommen.

Dennoch liegen die Vorzüge, die ein Social Bookmarking-Dienst für den User schafft, klar auf der Hand: Die gespeicherten Bookmarks/Artikel/Dateien sind von jedem Rechner aus jederzeit erreichbar und darüber hinaus ggf. auch für die Allgemeinheit von Nutzen. Die Community dient zudem als Filter für die Bookmarks, um Inhalte mit geringerer Qualität auszusortieren. Thematisch können die gefundenen Treffer mehr in die Tiefe gehen, als bei algorithmischen Suchmaschinen. Auch können sich Gruppen einen Account mit Bookmarks teilen, die für ihre jeweiligen Mitglieder relevant sind (Vereine, Arbeitskollegen, Studenten).

3.2 Frage-Antwort-Dienste

Bei Frage-Antwort-Diensten handelt es sich ebenfalls um eine spezielle Ausprägungsform sozialer Softwarelösungen. Sie bieten dem Anwender die Möglichkeit, Fragen an andere Teilnehmer dieser Dienste zu stellen und beantworten zu lassen. Suchanfragen werden somit in einem von Menschen erzeugten Datenbestand durchgeführt, also von einer Community von Freiwilligen beantwortet. Dem liegt die Annahme zugrunde, dass algorithmenbasierte Suchmaschinen bestimmte Anfragen nicht ohne weiteres beantworten können. Exemplarisch hierfür seien z. B. Meinungsfragen oder Fragen nach Ratschlägen genannt. Als Anreiz zur Beantwortung von Fragen erhalten die Anwender in der Regel Punkte, die mit einem Rangsystem verknüpft sind (z. B. bei Lycos iQ). Je mehr Fragen ein Anwender beantwortet, desto höher kann er in seinem Rang aufsteigen und einen gewissen Status in der Community aufbauen. In technischer Hinsicht liegt den meisten Frage-Antwort-Diensten – analog den sozialen Bookmarksystemen – das Prinzip des Tagging zugrunde. Dies ermöglicht es z. B., Fragen in Themengebiete zusammenzufassen oder Anwender mit ähnlichen Interessen zu identifizieren, weist allerdings auch dieselben oben beschriebenen Schwachstellen auf (fehlerhafte Tags, unsachgemäße Zuordnungen usw.).

Dem Nutzer steht mittlerweile ein breites Angebot an Frage-Antwort-Diensten zur Verfügung, die sich in ihrer Funktionalität nicht wesentlich voneinander unterscheiden. Wichtiger als technische Features ist jedoch die Größe der Community. Nur wenn eine Frage in angemessener Zeit und zur Zufriedenheit des Fragestellers beantwortet wird, können solche Dienste einen Nutzen stiften. Je größer die Community dabei ausfällt, desto höher ist auch die Wahrscheinlichkeit, eine kompetente Antwort zu erhalten. In Deutschland spielen vor diesem Hintergrund

lediglich die Suchdienste Yahoo! Clever, Lycos iQ und guteFrage.net eine wichtige Rolle auf dem Markt.

Gemessen an ihrem wirtschaftlichen Erfolg kann gegenwärtig keiner der hier skizzierten sozialen Suchdienste überzeugen und einem Vergleich mit algorithmenbasierten Suchmaschinen standhalten. Auffallend ist, dass sich der Suchmaschinenriese Google aus diesem Geschäftsfeld komplett zurückgezogen hat; bis vor kurzer Zeit hat z. B. auch Google den kostenpflichtigen Frage-Antwort-Dienst Google Answers betrieben.

3.3 Semantische Suchdienste

Neben sozialen Suchdiensten werden weiterhin semantische Suchmaschinen als Alternative für algorithmenbasierte Suchmaschinen diskutiert. Um für deren Funktionsweise und Bedeutung genauer zu verstehen, ist es wichtig, zunächst auf die Vision des semantischen Webs im Allgemeinen einzugehen, die von Tim Berners-Lee geprägt wurde. Danach sollen Mensch und Computer ein gemeinsames Verständnis über einen bestimmten Ausschnitt der Realität haben und Webseiten inhaltlich interpretieren können. Zu diesem Zweck entwickelte Berners-Lee den Vorschlag, zusätzlich zu HTML, der Beschreibungssprache für die Ausgabe einer Webseite, einen Standard zur Beschreibung der Bedeutung der verwendeten Inhalte zu entwickeln und zu verwenden [13]. Diese Sprache würde es automatischen Verfahren und Computern erleichtern, die Inhalte einer Webseite zu verstehen und sinnvolle Schlussfolgerungen daraus zu ziehen.

Eine grundsätzliche Möglichkeit, die Bedeutung von Begriffen und deren Beziehungen untereinander zu modellieren sind Ontologien. Dabei handelt es sich um Datenstrukturen, in denen Begriffe als Konzepte und ihre Beziehungen als Relationen abgebildet werden. Würden Webseiten mit solcherart strukturierter Zusatzinformation versehen, wäre es z. B. möglich, mit automatischen Methoden zwischen gleichlautenden Begriffen mit unterschiedlicher Bedeutung zu unterscheiden (z. B. Golf als Auto, Sport oder Meeresbucht) oder Informationen zu ähnlichen Themen zu finden (z. B. von Golf zu Cricket als Sportart, die ebenfalls mit einem Schläger auf einem Rasen ausgeführt wird). Somit würden sich bei der Websuche viele Möglichkeiten eröffnen, Suchanfragen zu verstehen bzw. genauer zu formulieren.

Dabei stellt sich natürlich die Frage, wer diese Metainformationen zu den Webseiten erstellt und auf Korrektheit überprüft. Ontologien sind zwar ein mächtiges Hilfsmittel, stellen sich aber auch als Hindernis dar, wenn die Annotationen von durchschnittlichen Internetnutzern erstellt werden sollen. Hierzu sind Ontologien zu komplex und für den Laien nur schwer verständlich. Ein Ausweg liegt in der automatischen Erstellung von Ontologien. Dazu wird versucht, mit unterschiedlichen Methoden aus den Gebieten Text Mining, Informationsextrahierung und maschinelles Lernen, den Inhalt einer Seite automatisch zu erfassen, in Ontologiestrukturen abzulegen und mit bereits vorhandenen Ontologien zu verknüpfen. Allerdings kann man sich leicht vorstellen, dass auch dieser Ansatz sehr anspruchsvoll und daher noch nicht so ausgereift ist, dass es bereits darauf basierende verlässliche und robuste Produkte gäbe.

Insgesamt muss man also feststellen, dass das semantische Web als zusätzliche Bedeutungsschicht, die das World Wide Web vollständig umfasst, eine Vision ist, die vermutlich so nicht erreichbar sein wird. Ein Beispiel dafür ist der in der Presse öfters

als „Google-Killer“ gehandelte semantische Suchdienst Powerset, der trotz einer relativ langen Entwicklungsphase noch immer nicht wie geplant online ist und sich derzeit noch als Wikipedia-Suchmaschine beschränkt; und auch hier sind die Ergebnisse noch nicht zufrieden stellend.

Dennoch ist es beachtlich, welche Fortschritte in der Forschung bereits erzielt wurden. Zieht man dazu noch das dem Normalnutzer zugänglichere Tagging zur Erzeugung von Metainformationen in Betracht, so lassen sich Vorgehensweisen vorstellen, die aus den zwar unstrukturierten und fehlerbehafteten, aber von menschlichen Benutzern bereitwillig erstellten Zusatzinformationen in Kombination mit bereits vorhanden Ontologien und semantischen Verfahren die gewünschten Annotationen erzeugen.

Eine andere Möglichkeit zur Verbreitung semantischer Annotationen, die sich derzeit immer größerer Verbreitung erfreut, ist die Verwendung so genannter Mikroformate im HTML-Code der Webseite. Hierdurch bekommen die Entwickler von Webseiten die Möglichkeit, bestimmte Informationen in Standardformaten direkt als unsichtbare Informationen auf der Webseite zu hinterlegen. Sobald solche Mikroformate von den großen Suchmaschinenbetreibern unterstützt werden (wie kürzlich von Yahoo! Angekündigt [14]) und dadurch die Aussicht besteht, dass die gesuchte Information von den Standardsuchmaschinen besser gefunden wird, werden weitere Seitenbetreiber Mikroformate einsetzen. Eine breite Toolunterstützung, die das Erstellen von Mikroformaten erleichtert oder gar automatisiert, würde wohl bald folgen. Ist so ein Punkt erreicht, könnte zumindest ein Teil der Semantic-Web-Vision Realität werden.

3.4 Konvergente Suchdienste

Neben den bislang skizzierten Suchdiensten existieren noch weitere, die Eigenschaften mehrerer Suchdienste vereinen und als konvergent bezeichnet werden (vgl. Abb. 1.1). Sie verfolgen generell das Ziel, die Vorteile der bislang nur isoliert voneinander fungierenden Suchdienste miteinander zu verknüpfen, um die Qualität der Suchergebnisse zu verbessern oder zumindest deren Nachteile zu mindern. Ein Beispiel hierzu ist die Suchmaschine Wikia Search, die vom Wikipedia-Gründer Jimmy Wales frei zur Verfügung gestellt wird und das Wikipedia-Prinzip des freien Editierens mit den Eigenschaften einer algorithmischen Suchmaschine verbinden soll. Das bedeutet, dass mittels Crawling ein Index erstellt wird, der dann von den Nutzern bearbeitet werden soll. Die Suchergebnisse des im Januar 2008 eingeführten Dienstes sind wie erwartet noch nicht vergleichbar mit anderen Suchmaschinen und erst die nächste Zeit wird zeigen, ob aus diesem Dienst ein ähnlich erfolgreiches Projekt wie Wikipedia entstehen wird. Weitere Beispiele für die konvergente Suche sind die personalisierte Suche und die implizite Suche in so genannten Clickstreams. So soll der derzeit in der Betaphase befindliche Suchdienst Twine. Komponenten sozialer Netzwerke und semantischer Suche in einer persönlich konfigurierbaren Suche zusammenführen [15, 16]. Basierend auf einer Semantic-Web-Datenstruktur, die es ermöglicht, Inhalte computerverständlich abzulegen, soll der Nutzer dabei unterstützt werden, seine gefundenen Informationen mit anderen auszutauschen und selber wiederum durch den Austausch mit anderen besser an gesuchte Informationen zu gelangen. Bei Clickstreams hingegen werden die impliziten Informationen ausgenutzt, die der Nutzer durch sein Surfverhalten hinterlässt. Klicks auf Links und

Seitenverweildauern werden mit Methoden des Data Mining und des maschinellen Lernens ausgewertet, um daraus zu folgern, was der Nutzer „eigentlich“ suchte, ohne eine explizite Suchanfrage gestellt zu haben. Da solche Verfahren allerdings nur mit Hilfe der Auswertung persönlicher Daten auf den Systemen der Internetprovider funktionieren können, stößt dieser Ansatz in der öffentlichen Diskussion auf breite Kritik.

4. FAZIT

Jeder der in diesem Beitrag untersuchten Suchdienste birgt Vor- und Nachteile, die für eine Empfehlung, welcher Dienst bei welcher Art von Suche herangezogen werden sollte oder wie die Ansätze ggf. zu kombinieren sind, miteinbezogen werden müssen. Die algorithmische Suchmaschine bietet dem Nutzer mit aktuellen Meldungen und News eine Aktualität, die andere Dienste in diesem Umfang nicht bieten. Eine Suche nach variablen Datenformaten und verschiedenen Kategorien wie Nachrichten, Karten, Shops und Bildern leistet in vollem Umfang ebenfalls nur die algorithmische Suche. Weder Frage-Antwort- noch Social Bookmarking-Dienste generieren Suchergebnisse in dieser Menge unter Abdeckung aller Themenbereiche. Diese große Menge der Suchergebnisse ist aber gleichzeitig ein Nachteil, da nicht alle Treffer für die Suchanfrage relevant sind. Ebenso sind hier das Problem des Suchmaschinen-Spamming und die Schwierigkeiten durch die Uneindeutigkeit menschlicher Sprache zu bedenken. Der Nutzer muss also durch eine große Anzahl von Treffern bewerten, um das für ihn relevante Ergebnis zu finden. Dabei enthält der Index einer algorithmischen Suchmaschine noch nicht einmal ansatzweise alle im WWW enthaltenen Webseiten und Informationen.

Eine mögliche Lösung für diese Probleme könnte in konvergenten Diensten liegen – einer Kombination einer algorithmischen Vorgehensweise mit einem Social Bookmarking-Dienst, da über eine Anfrage durch einen Social Bookmarking-Dienst auch speziellere Webseiten gefunden und durch die Community bewertet werden können. Somit bieten Social Bookmarking-Dienste die kollektive Intelligenz der Community zur Steigerung der Qualität der Suchergebnisse. Nachteilig beim Social Bookmarking-Dienst sind wiederum die fehlende Aktualität und eventuelle thematische Lücken, da hier ja nicht automatisiert aktualisiert und indexiert wird. Dennoch ist Social Bookmarking eine interessante Variante, tiefer ins Web einzutauchen.

Durch die Kombination mit einem Frage-Antwort-Dienst könnte dem User ein zusätzlicher Service geboten werden. Die Stärke der Frage-Antwort-Dienste liegt in der Beantwortung von gezielt gestellten individuellen Fragen. Fragen im Erfahrungsbereich können weder von einer algorithmischen Suchmaschine noch von einem Social Bookmarking-Dienst beantwortet werden. Die Richtigkeit der Antwort kann ebenfalls nicht garantiert werden. Auch wird es hierbei immer Lücken in den behandelten Themenbereichen geben, da nur Antworten auf Fragen gegeben werden können, die von Nutzern gestellt wurden. Unter Umständen muss der Nutzer dann eine längere Zeit auf die Beantwortung seiner Frage warten. Durch den starken Community-Charakter der Frage-Antwort-Dienste liegt der Schwerpunkt weniger auf wissenschaftlichen Themen, sondern eher auf Themengebieten wie Freizeit, Kultur

und Medien. Dadurch haben Frage-Antwort-Dienste oft eher einen Chat- oder Forencharakter.

Zusammenfassend betrachtet kann man jedoch feststellen, dass sich trotz aller offensichtlichen Herausforderungen schon bei dieser ersten Betrachtung der existierenden Suchdienste und ihrer prinzipiellen Kombinationsmöglichkeiten ein Bild ergibt, das den Versuch einer Integration insbesondere sozialer Suchansätze in algorithmische Vorgehensweisen als lohnenswert erscheinen lassen. Es soll daher in diesem vorliegenden Band detailliert untersucht werden, wie die betrachteten Dienste genau funktionieren, welche Erfahrungen damit gemacht wurden, welche Qualitätsunterschiede sich messen lassen, und ob eine Kombination tatsächlich sinnvoll scheint.

Verwendete Literatur

1. Gasser, Urs/Thurman, James (2007): Themen und Herausforderungen der Regulierung von Suchmaschinen. In: Machill, Marcel/Beiler, Markus (Hrsg.): Die Macht der Suchmaschinen, Köln 2007, S. 44-64.
2. Die Grünen (2005): Suchmaschinen: Das Tor zum Netz. Onlinedokument: <http://www.gruene-bundestag.de/cms/publikationen/dokbin/63/63265.pdf> [Abruf am 14.05.2008]
3. Die Grünen (2006): Google verleibt sich YouTube ein – und wird noch mächtiger. Onlinedokument: http://www.gruene-bundestag.de/cms/presse/dok/150/150791.google_verleibt_sich_youtube_ein_und_wir.html [Abruf am 14.05.2008]
4. Gray, Matthew (1996): Internet Statistics. Growth and Usage of the Web and the Internet. Onlinedokument: <http://www.mit.edu/~mkgray/net/> [Abruf am 24.05.2008]
5. ComScore (2008): Suchabfragen im April 2008. Onlinedokument: www.comscore.com [Abruf am 01.06.2008]
6. Schäfer, Karim (2006): Medienbranche mit guten Perspektiven bis 2010: Digitalisierung schafft robustes Wachstum. In: Innovations Report, 21.06.2006.. Onlinedokument: http://www.innovations-report.de/html/berichte/kommunikation_medien/bericht-66578.html [Abruf am 14.05.2008]
7. Microsoft Advertising (2008): Das Internet entwickelt sich bis 2010 zum weltweit drittgrößten Medium. Onlinedokument: http://advertising.microsoft.com/deutschland/Branchen-News?Adv_NewsID=891 [Anruf am 14.05.2008]
8. OVK (2007): Online-Report 2007/02. Onlinedokument: http://www.bvdw.org/fileadmin/downloads/marktzahlen/basispraesentationen/OVK_OnlineReport_2007_02.pdf [Abruf am 14. 05 2008]
9. Welt Online (2007): Google knackt die Milliarden-Marke. In: Welt Online, 18.10.2007. Onlinedokument: http://www.welt.de/wirtschaft/article1278520/Google_steigert_Gewinn_um_46_Prozent.html [Abruf am 24.05.2008]
10. Komus, Ayelt (2006): Social Software als organisatorisches Phänomen – Einsatzmöglichkeiten im Unternehmen. In: Hildebrand, Knut/Hofman, Josephine (Hrsg.): Social Software. HMD – Praxis der Wirtschaftsinformatik, H. 252, S. 36-14.
11. Gräfe, Gernot/Maaß, Christian./Heß, Andreas (2007): Alternative Search Engines: Seven Hypotheses Regarding the Future of Social Bookmarking. In: Software, Agents and Services for Business, Research and E-Sciences - Conference on Social Semantic Web (SABRE/CSSW 2007), Leipzig 2007.

12 Status quo und Entwicklungstendenzen

12. Bächle, Michael (2006): Social Software. In: Informatik Spektrum Bd. 29, H. 2, S. 121-124.
13. Berners-Lee, Tim/ Hendler, James/ Lassila, Ora (2001). The Semantic Web. In: Scientific American Magazine, Mai 2001. Onlinedokument: <http://www.sciam.com/article.cfm?id=the-semantic-web&ref=sciam> [Abgerufen am 18.03.2008]
14. Kumar, Amit (2008): The Yahoo! Search Open Ecosystem. In: Yahoo! Search Blog, 13.03.2008. Onlinedokument: <http://www.ysearchblog.com/archives/000527.html> [Abruf am 18. 03. 2008]
15. MacManus, Richard (2007): Twine: The First Mainstream Semantic Web App? In: ReadWriteWeb, 18.10.2007. Onlinedokument, http://www.readriteweb.com/archives/twine_first_mainstream_semantic_web_app.php [Abruf am 26.03.2008]
16. Schonfeld, Erick (2007): Twine Launches A Smarter Way To Organize Your Online Life. In: TechCrunch, 19.10.2007. Onlinedokument: <http://www.techcrunch.com/2007/10/19/twine-launches-a-smarter-way-to-organize-your-online-life/> [Abruf am 26.03.2008]